(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 206 104 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
15.05.2002 Bulletin 2002/20

(51) Int Cl.7: **H04M 3/32**

(21) Application number: **00203936.0**

(22) Date of filing: **09.11.2000**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(71) Applicant: **Koninklijke KPN N.V.**
9726 AE Groningen (NL)

(72) Inventors:
- **Appel, Symon Ronald**
  2627 AL Delft (NL)
- **Beerends, John Gerard**
  4585 PB Hengstdijk (NL)
- **Hekstra, Andries Pieter**
  2252 KM Voorschoten (NL)

(74) Representative: **Kruk, Wiggert Johan et al**
Koninklijke KPN N.V.,
Intellectual Property Dep.,
P.O.Box 95321
2509 CH Den Haag (NL)

(54) Measuring a talking quality of a telephone link in a telecommunications network

(57) For measuring the influence of noise on the talking quality of a telephone link in a telecommunications network, a talker speech signal (s(t)) and a degraded speech signal (s'(t)) are fed to an objective measurement device (32) for obtaining an output signal (q) representing an estimated value of the talking quality. The degraded signal includes a returned signal (r(t)) originating from the network during transmission of the talker speech signal over the telephone link. The objective measurement carried out by the device is a modified PSQM-like measurement, which is modified as to include a modelling (32b) of masking effects in consequence of noise present in the returned signal. Preferably the modelling includes a noise suppression (42) carried out to a difference signal (D(t,f)) in the loudness density domain using a noise estimation (41).

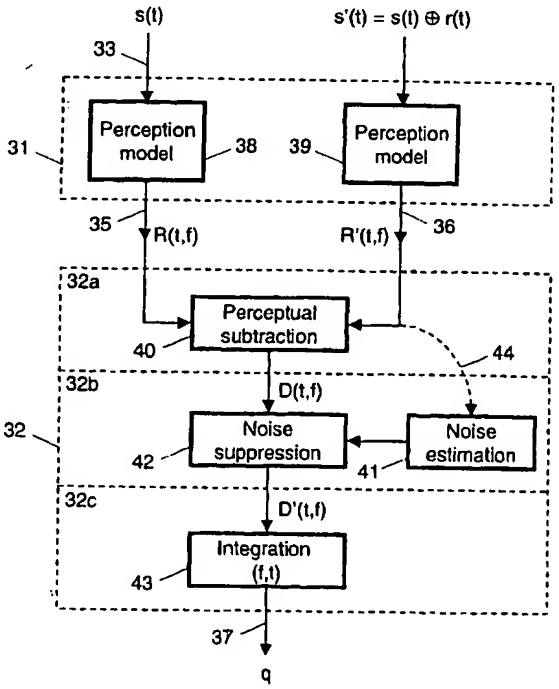

FIG. 3

EP 1 206 104 A1

Description

A. BACKGROUND OF THE INVENTION

[0001] The invention lies in the area of measuring the quality of telephone links in telecommunications systems. More in particular, it concerns measuring a talking quality of a telephone link in a telecommunication network, i.e. measuring the influence of returned signals such as echo disturbances and side tone distortions on the perceptual quality of a telephone link in a telecommunications system as subjectively observed by a talker during a telephone call.

[0002] Such a method and a corresponding device are described in the not timely published international patent application PCT/EP00/08884 (Reference [1]; for more bibliographical details relating to the references, see below under D.), which is incorporated by reference in the present application. According to the described method and device for measuring the influence of echo on the perceptual quality on the talker's side of a telephone link in a telecommunications network, a talker speech signal and a combined signal are fed to an objective measurement device, such as a PSQM system, for obtaining an output signal representing an estimated value of the perceptual talking quality. The combined signal is a signal combination of a returned signal originating from the network and corresponding to the talker speech signal, and the talker speech signal itself. The described technique has the following problem. In case the returned signal contains signal components not directly related to the voice of the talker, like noise present in the telephone system, noise derived from the background noise of the talker at the other side of the telephone connection, or noise derived from interfering signals, such signal components may have a so-called masking effect on the echo, which then results in an increase of the subjectively perceived talking quality. Objective measurement systems such as based on the Perceptual Speech Quality Measurement (PSQM) model, recommended by the ITU-T Recommendation P.861 (see Reference [2]), however, will interpret noise components generally in terms of a decrease in quality. This problem may be tried to be solved by using noise suppression techniques as generally known in the world of speech processing (see e.g. References [3],-,[6]). However, these known suppression techniques are developed for optimising listening quality, and are not suited for the measurement and optimisation of talking quality. Talking quality differs from listening quality, especially in the effect of masking noise and masking by one's own voice. Noise in general decreases listening quality but increases talking quality.

B. SUMMARY OF THE INVENTION

[0003] The main object of the present invention is to provide for an improved objective measurement method and corresponding device for measuring a talking quality of a telephone link in a telecommunication network, i.e. for measuring the influence of returned signals such as echo, side tone distortion, inclusive the influence of noise, on the perceptual quality on the talker's side of the telephone link, which do not possess said problem.

[0004] A method for measuring a talking quality of a telephone link in a telecommunications network according to the preamble of claim 1, as described in Reference [1], is, according to the invention, characterised as in claim 1.

[0005] A device for measuring a talking quality of a telephone link in a telecommunications network according to the preamble of claim 10, as described in reference [1], is, according to the invention, characterised as in claim 10.

[0006] The invention is based on the appreciation that objective measurement systems such as PSQM, and which are covered by the above-mentioned Recommendation P.861, have been developed for measuring the listening quality of speech signals. Therefore, in order to provide a similar objective measurement for measuring the talking quality of a telephone link, the step of modelling echo masking effects is introduced in the objective measurement method and device.

[0007] According to the Recommendation P.861 at first a speech signal, which is an output signal of an audio- or speech processing or transporting system, and of which the signal quality has to be assessed, and a reference signal are mapped to representation signals of a psycho-physical perception model of the human auditory system. These representation signals are in fact the compressed loudness density functions of the speech and reference signals. Then two operations, which imply an asymmetry processing and a silent interval weighting in order to model two cognitive effects, are carried out on a difference signal of the two representation signals in order to produce the quality signal which is a measure for the auditory perception of the speech signal to be assessed. However, it is known that noise in the echo signal, especially background noise originating at the side of the B subscriber of the telephone link, can have a masking effect on the echo signal, thus leading to an improvement of the subjectively perceived talking quality. Then it was realised that in the operations carried out on the difference in the algorithm of the Recommendation P.861 noise in the echo signal will be interpreted as an introduced distortion, leading to a deterioration of the objectively measured talking quality, and therefore these operations should be modified and/or supplemented by a step of modelling echo masking effects of noise.

[0008] Therefore a preferred embodiment of the method and of the device of the present invention are characterised according to claim 2 and claim 11, respectively.

[0009] Further preferred embodiments of the method and the device of the invention are summarised in the

various subclaims.

C. REFERENCES

[0010]

[1] PCT/EP00/08884 (of applicant; filing date: 08.09.2000);
[2] ITU-T recommendation P.861: Objective quality measurement of telephone band (330-3400 Hz) speech codecs, August 1996;
[3] R. Le Bouquin, " Enhancement of Noisy Speech Signals: Applications to Mobile Radio Communications", Speech Communication, vol. 18, pp. 3-19 (1996);
[4] J.-H Chen and A. Gersho, "Adaptive Postfiltering for Quality Enhancement of Coded Speech", IEEE Trans. on Speech and Audio Processing., vol. 3, pp. 59-71 (1995 Jan);
[5] D. E. Tsoukalas, J. Mourjopoulos and G. Kokkinakis, " Perceptual Filters for Audio Signal Enhancement", J. Audio Eng. Soc., vol. 45, pp. 22-36 (1997 Jan/Feb);
[6] F. Xie and D. van Compernolle, "Speech Enhancement by Spectral Magnitude Estimation - A unifying Approach", Speech Communication, vol. 19, pp. 89-104 (1996).

[0011] All references are considered to be incorporated into the present application.

D. BRIEF DESCRIPTION OF THE DRAWING

[0012] The invention will be further explained by means of the description of exemplary embodiments, reference being made to a drawing comprising the following figures:

FIG. 1 schematically shows an example of a usual telephone link in a telecommunications network;
FIG. 2 schematically shows an earlier described set-up for measuring a talking quality of a telephone link using a known objective measurement technique for measuring a perceptual quality of speech signals;
FIG. 3 schematically shows a device for an objective measurement of a talking quality of a telephone link according to the invention to be used in the set-up of FIG. 2;
FIG. 4 shows a flow diagram of the detailed operation of a part of the device shown in FIG. 3;
FIG. 5 schematically shows a modification in a further part of the device shown in FIG. 3.

E. DESCRIPTION OF EXEMPLARY EMBODIMENTS

[0013] Delay and echo play an increasing role in the quality of telephony services because modern wireless and/or packet based network techniques, like GSM, UMTS, DECT, IP and ATM inherently introduce more delay than the classical circuit switching network techniques like SDH and PDH. Delay and echo together with the side tone determine how a talker perceives his own voice in a telephone link. The quality with which he perceives his own voice is defined as the talking quality. It should be distinguished from the listening quality which deals with how a listener perceives other voices (and music). Talking and listening quality together with the interaction quality determine the conversational quality of a telephone link. Interaction quality is defined as the ease of interacting with the other party in a telephone call, dominated by the delay in the system and the way it copes with double talk situations. The present invention is related to the objective measurement of talking quality of a telephone link, and more particular to account for the influence of noise therein.

[0014] FIG. 1 schematically shows an example of a usual telephone link established between an A subscriber and a B subscriber of a telecommunications network 10. Telephone sets 11 and 12 of the A subscriber and the B subscriber, respectively, are connected by way of two-wire connections 13 and 14 and four-wire interfaces, namely, hybrids 15 and 16, to the network 10. Through the network, the established telephone link has a forward channel including a two-wire part, i.c. two-wire connections 13 and 14, and a four-wire send part 17, over which speech signals from the A subscriber are conducted, and a return channel including a two-wire part, i.c. two-wire connections 14 and 13, and a four-wire receive part 18, over which speech signals from the B subscriber are conducted. A speech signal s striking the microphone M of the telephone set 11 of the A subscriber, is passed on, by way of the forward channel (13, 17, 14) of the telephone link, to the earphone R of telephone set 12, and becomes audible there for the B subscriber as a speech signal s" affected by the network. Each speech signal s(t) on the forward channel generally causes a returned signal r(t) which, particularly due to the presence of said hybrids, includes an electrical type of echo signal on the return channel (18, 13) of the telephone link, and this is passed on to the earphone R of the telephone set 11, and may therefore disturb the A subscriber there. Furthermore the acoustic and/or mechanical coupling of the earphone or loudspeaker signal to the microphone of the telephone set of the B subscriber may cause an acoustic type of echo signal back to the telephone set of the A subscriber, which contributes to the returned signal. In an end-to-end digital telephone link (such as in a GSM system or in a Voice-over-IP system) such acoustic echo signal is the only type of echo signal that contributes to the return signal.
Summarizing a returned signal r(t) may include, at various stages in the return channel of a telephone link as caused by a speech signal s(t) in the forward channel of the telephone link:

- a signal r1 representing acoustic echo;
- a signal r2 representing an electrical echo possibly in combination with the acoustic echo;
- a signal r3 which represents the signal r2 as affected, i.e. delayed or distorted, by the network 10;
- a signal r4 which represents the signal r3 in combination with a side tone signal, and
- a signal r5 which is an acoustic signal derived from the signal r4, that also includes the locally generated side tone.

[0015] FIG. 2 shows schematically a set-up for measuring a talking quality of a telephone link using a known objective measurement technique for measuring a perceptual quality of speech signals, as described in reference [1]. The set-up comprises a system or telecommunications network under test 20, hereinafter for briefness' sake referred to as network 20, and a system 22 for the perceptual analysis of speech signals offered, hereinafter for briefness' sake only designated as PSQM system 22. Any talker speech signal s(t) is used, on the one hand, as an input signal of the network 20 and, on the other hand, as first input (or reference) signal of the PSQM system 22. A returned signal r(t) obtained from the network 20, which corresponds to the input talker speech signal s(t), is combined, in a combination circuit 24, with the talker speech signal s(t) to provide a combined speech signal s'(t), which is then used as a second input (or degraded) signal of the PSQM system. If necessary, the signal s(t) is scaled to the correct level before being combined with the returned signal r(t) in the combination circuit. An output signal q of the PSQM system 22 represents an estimate of the talking quality, i.e. of the perceptual quality of the telephone link through the network 20 as it is experienced by the telephone user during talking on his own telephone set. Here use may be made of signals stored on data bases. These signals may be obtained or have been obtained by simulation or from a telephone set (e.g. signal r4 in the electrical domain or signal r5 in the acoustic domain) of the A subscriber in the event of an established link during speech silence of the B subscriber. The two-wire connection between the telephone subscriber access point and the four-wire interface with the network does not, or hardly, contribute to the echo component in the returned signal r(t) (of course, it does contribute to the echo component in a returned signal occurring in the return channel of the B subscriber of the telephone link). However, any such signal contribution has a short delay and, as a matter of fact, forms part of the side tone.

[0016] The signals s(t) and r(t) may also be tapped off from a four-wire part 17 of the forward channel and the four-wire part 18 of the return channel near the four-wire interface 15, respectively. This offers, as already described in reference [1], the opportunity of a permanent measurement of the talking quality in the event of established telephone links, using live traffic non-intrusively.

[0017] The system or network being tested may of course also be a simulation system, which simulates a telecommunications network.

[0018] The described technique has, however, the following problem. Since a system or network under test generally will not be ideal, any returned signal r(t) will contain also signal components not directly related to the voice of the talker, like noise present in the telephone system, noise derived from the background noise of the listener at the other side of the telephone connection, or noise derived from interfering signals. In such a case these signal components may have a so-called masking effect on the echo, which then results in an increase of the talking quality. Objective measurement systems like PSQM, however, which up to now have been developed for assessing the listening quality of speech signals, will interpret such noise components in terms of a decrease in quality. In the following a method and a device are described which in essence imply a modification of a PSQM-like algorithm as recommended by the ITU-T Recommendation P.861, in order to avoid the problem and to make the existing algorithm suitable for objectively measuring the talking quality with a higher correlation with a subjectively measured talking quality, when used in a set-up as shown in FIG. 2, than without the modification.

[0019] FIG. 3 shows schematically a measuring device for objectively measuring the perceptual quality of an audible signal. The device comprises a signal processor 31 and a combining arrangement 32. The signal processor is provided with signal inputs 33 and 34, and with signal outputs 35 and 36 coupled to corresponding signal inputs of the combining arrangement 36. A signal output 37 of the combining arrangement 36 is at the same time the signal output of the measuring device. The signal processor includes perception modelling means 38 and 39, respectively coupled to the signal inputs 33 and 34, for processing input signals s(t) and s'(t) and generating representation signals R(t,f) and R'(t,f) which form time/frequency representations of the input signals s(t) and s'(t), respectively, according to a perception model of the human auditory system. The representation signals are functions of time and frequency (Hz scale or Bark scale). The signal processing, as usual, is carried out frame-wise, i.e. the speech signals are split up in frames that are about equal to the window of the human ear (between 10 and 100ms) and the loudness per frame is calculated on the basis of the perception model. Only for reasons of simplicity this frame-wise processing is not indicated in the figures.

[0020] The representation signals R(t,f) and R'(t,f) are passed to the combining arrangement 32 via the signal outputs 35 and 36. In the combining arrangement of the known PSQM-like algorithm at first a difference signal of the representation signals is determined followed by various processing steps carried out on the difference signal. The last ones of the various processing steps imply integration steps over frequency and time resulting

in a quality signal q available at the signal output 37.

[0021] In case of determining a listening quality the input signal s'(t) is an output signal of an audio- or speech signals processing or transporting system, of which the signal processing or transporting operation is assessed, while the input signal s(t), being the corresponding input signal of the system to be assessed, is used as reference signal. For determining a talking quality, however, where, as described with reference to FIG. 2, the input signal s'(t) is a combination of the signal s (t) and the returned signal r(t), the known combining arrangement should be modified.

[0022] According to the recommended PSQM-like algorithm (see reference [2], more particularly FIGURE 3/P.861) the various processing steps carried out by (within) the combining arrangement, include asymmetry processing and silent interval weighting steps for modelling some perceptual effects. It is known that noise in the echo signal, especially background noise originating at the side of the B subscriber of the telephone link, has a masking effect on the echo signal, thus leading to an improvement of the subjectively perceived talking quality. Then it was realized that the presence of the steps for modelling the cognitive effects in the algorithm, however, in which noise in the echo signal will be interpreted as an introduced distortion, would lead to a deterioration of the objectively measured talking quality, and therefore could not be maintained as such.

[0023] Instead, for correctly measuring the talking quality, a step of modelling masking effects which noise present in the returned signal could have on perceived echo disturbances, is introduced. Such a modelling step could be based on a possible separation of echo components and noise components present in the returned signal r(t). However a reliable modelling could be reached in a different, simpler manner. This modelling step implies a specific noise suppression step carried out on the difference signal by using an estimated value for the noise. Therefore the combining arrangement 32 comprises:

- in a first part 32a, a subtraction means 40 for perceptually subtracting the two representation signals R(t,f) and R'(t,f) received from the signal processor 31 and generating a difference signal D(t,f),
- in a second part 32b, a noise estimating means 41 for generating an estimated noise value Ne for the noise present in the input signal s'(t), and a noise suppression means 42 for deriving from the difference signal D(t,f) and the estimated noise value Ne a modified difference signal D'(t,f), and
- in a third part 32c, integration means 43 for integrating the modified difference signal D'(t.f) successively to frequency and time and generating the quality signal q.

[0024] The estimated noise value Ne may be a predetermined value, e.g. derived from the type of telephone link, or is preferably obtained from one of the representation signals, i.c. R'(t,f), which is visualised in FIG. 3 by means of a broken dashed line between the signal output 36 with a signal input 44 of the noise estimation means 41. The representation signals R(t,f) and R'(t,f) are as usual loudness density functions of the reference and degraded speech signals s(t) and s'(t), respectively. The output signal of the subtraction means 40, i.c. D(t, f), represents the signed difference between the loudness densities of the degraded (i.e. distorted by the presence of echo, side tone and noise signals in the returned signal) and the reference signal (i.e. the original talker speech signal), preferably reduced by a small perceptual correction, i.c. a small density correction for so-called internal noise.

[0025] The resulting difference signal D(t,f), which is in fact a loudness density function, is subjected to a background masking noise estimation. The key idea behind this is that, because talkers during a telephone call will always have silent intervals in their speech, during such intervals (of course after the echo delay time) the minimum loudness of the degraded signal over time is almost completely caused by the background noise. Since the speech signal processing is carried out in frames, this minimum may be put equal to a minimum loudness density Ne found in the frames of the representation signal R'(t,f). This minimum Ne can then be used to define a threshold value T(Ne) for setting the content of all frames of the difference signal D(t,f), that have a loudness below this threshold, to zero, leaving the content of the other frames unchanged. The set-to-zero frames and the unchanged frames together constitute a signal from which the modified difference signal D'(t,f), the output signal of the noise suppression means 42, is derived (see below). Consequently, the standard Hoth noise background masking noise, used in the main step of the PSQM-like algorithm of deriving the representation signals, has to be omitted from the algorithm.

[0026] FIG. 4 shows schematically by means of a flow diagram more in detail the modelling step as carried out on the difference signal D(t,f) by the noise suppression means 42 using the estimated noise value Ne produced by the noise estimating means 41. Again it is emphasized that, although for sake of simplicity only not indicated in the figures, the signal processing is understood to be frame-wise. The flow diagram includes the following boxes:

- box 45 indicating a step of integrating the representation signal R'(t,f), as produced by the signal processor 31 via output 36, over frequency, resulting in a loudness degraded signal R'(t);
- box 46 indicating a step of determining the estimated noise value Ne for the noise present in the loudness degraded signal R'(t), Ne being equal to the minimum value of the loudness found in the loudness degraded signal R'(t) ;
- boxes 47, 48 and 49 indicating a step of subjecting

the difference signal D(t,f) to a criterion C by means of which from the difference signal a thresholded difference signal $D_c(t,f)$ is derived, box 48 indicating that $D_c(t,f) = D(t,f)$ for frames in which the loudness of the frames in the loudness degraded signal R'(t) suffices to the criterion and box 49 indicating that $D_c(t,f) = 0$ for frames in which the loudness of the frames in the loudness degraded signal R'(t) does not suffice to the criterion C;

- box 50 indicating a step of determining from the thresholded difference signal $D_c(t,f)$ the modified difference signal D'(t,f) by calculating a distortion loudness to signal loudness ratio (DSR) of the thresholded difference signal $D_c(t,f)$ and the loudness degraded signal R'(t), i.e. D'(t,f) = DSR(t,f).

[0027] Experimentally a suitable criterion C appeared to be that the loudness of the frames in the loudness degraded signal R'(t) is larger than or equal to the threshold value T(Ne) or not, choosing said threshold value to be a constant factor C_f times the estimated value Ne, i.e. $T(Ne) = C_f.Ne$. A suitable value for the constant factor appeared to be $C_f = 1.6$.

[0028] In calculating the DSR of the difference signal a clipping is carried out by introducing a threshold on the signal loudness, below which the signal loudness is set to that threshold. In an optimisation a threshold value of 4 Sone was found.

[0029] Finally the modified difference signal D' (t,f) is integrated by means of the integration means 43 at first over frequency using an Lp norm (i.e the generally known Lebesgue p-averaging function or Lebesgue p-norm) with p=0.8, and over time using an Lp norm with p=6, resulting in the output value q for the talking quality.

[0030] The quality output values of a thus modified objective measurement method and device for assessing the talking quality, as experimentally obtained for seven databases of test speech signals, showed high correlations (above 0.93) with the mean opinion scores (MOS) of the subjectively perceived talking quality.

[0031] For the measuring of the talking quality it is necessary that the representation signal R'(t,f) is a representation of the signal combination of the talker speech signal and the returned signal. To realise this, however, it is not necessary that the degraded signal s'(t) is a signal combination of these two signals as indicated in FIG. 2 (signal combinator 24) and in FIG. 3 ($s'(t)=s(t)\oplus r(t)$). It is also possible to use the returned signal (r(t)) as the degraded signal (s'(t)) and to obtain an intermediate signal in an intermediate stage of processing the reference signal, as carried out by the perception modelling means 38, which then is combined with a corresponding intermediate signal (Ps'(f)) obtained in a corresponding intermediate stage of processing the degraded signal, as carried out by the perception modelling means 39. Preferably the intermediate signal is a Fast Fourier Transform power representation (Ps(f)) of the reference speech signal (s(t)). This modification is shown schematically in FIG. 5 more in detail. The perceptual modelling means 38 and 39 carry out in a first stage of processing as usual (see reference [2]), respectively indicated by boxes 51 and 52, a step of determining a Hanning window (HW) followed by a step of determining a Fast Fourier Transform (FFT) power representation in order to produce the intermediate signals Ps(f) and Pr(f), which are FTT power representations of the talker speech signal s(t) and the degraded signal s'(t) which now equals the returned signal r(t), respectively. In a second stage of processing, respectively indicated by boxes 53 and 54, a step of frequency warping (FW) to pitch scale is carried out followed by steps of frequency smearing (FS) and intensity warping (IW), in order to produce the representation signals R(t,f) and R'(t,f). Between the first and second stages, as indicated by the boxes 52 and 54, an intermediate signal addition of the intermediate signals Ps(f) and Pr(f), indicated by signal adder 55, is carried out, the intermediate signal sum in addition being the input of the second processing stage (box 54). Before the intermediate signal addition can be applied, the intermediate signal P(s(f)) has to be scaled to the correct level as usual.

[0032] Consequently, when using such an intermediate signal addition ($Ps(f)\oplus Pr(f)$) inside the perception modelling means, instead of the external addition ($s'(t) = s(t)\oplus r(t)$), the combination circuit 24 becomes superfluous. In case a device as described with reference to FIG. 3, having included the modification as described with reference to FIG. 5, is used directly in a telephone link, in a way as already described in reference [1], then the input ports 33 and 34 of the device may be directly coupled to the four-wire parts 17 and 18 of the forward and return channel, respectively, of a telephone link.

Claims

1. Method for measuring a talking quality of a telephone link in a telecommunications network,
the method comprising a main step of subjecting a degraded speech signal s'(t) with respect to a reference speech signal s(t) to an objective measurement technique (32) for measuring a perceptual quality of speech signals, and producing a quality signal (q) which represents an estimated value concerning the talking quality, the reference speech signal being a talker speech signal (s(t)) and the degraded speech signal including a returned signal r (t), the returned signal being a signal which occurred or may occur in a return channel of the telephone link during the transmission of the talker speech signal in a forward channel of the telephone link,
characterised in that the main step is carried out by means of an objective measurement technique which includes a step of modelling masking effects in consequence of noise present in the returned sig-

nal.

2. Method according to claim 1, **characterised in that**

the main step comprises:

a first processing step of processing the degraded speech signal (s'(t)) and generating a first representation signal (R'(t,f)),
a second processing step of processing the talker speech signal (s(t)) and generating a second representation signal (R(t,f)), and
a combining step of combining the first and second representation signals as to produce said output signal (q),

the first representation signal (R'(t,f)) being a representation signal of a signal combination of the talker speech signal and the returned signal, and the combining step including said step of modelling masking effects in consequence of noise present in the returned signal.

3. Method according to claim 2, **characterised in that**

the combining step includes:

a step of subtracting (32a) the first representation signal from the second representation signal as to produce a difference signal (D(t,f)),
said step of modelling (32b) the masking effects of noise carried out on the difference signal as to produce a modified difference signal, and
a step of integrating (32c) the modified difference signal with respect to frequency and time as to produce the quality signal,

the modelling step including:

a first substep of producing (41) an estimated value (Ne) of the loudness of the noise present in the returned signal, and
a second substep of noise suppression (42; 46) carried out on the difference signal using said produced estimated value (Ne) as to produce the modified difference signal (D'(t,f)).

4. Method according to claim 3, **characterised in that** the second substep of noise suppression includes further substeps of:

deriving (46) from the estimated value (Ne) a loudness criterion (C),
setting (47, 48, 49) distorsions in the loudness (domain) of the difference signal, which do not suffice the criterion, to zero in the loudness (domain) of a thresholded difference signal (D_c(t, f)), and
deriving (50) the modified difference signal (D'(t,f)) by calculating a distorsion loudness to signal loudness ratio (DSR(t,f)) of the thresholded signal (D_c(t,f)) with respect to a loudness degraded signal (R'(t)) derived from the first representation signal (R'(t,f)).

5. Method according to any of the claim 2,-,4, **characterised in that** the estimated value of the noise loudness is derived from the first representation signal (R'(t,f)).

6. Method according to any of the claims 2,-,5, **characterised in that** the degraded signal (s'(t)) is a signal combination of the talker speech signal (s(t)) and the returned signal (r(t)).

7. Method according to any of the claims 2,-,5, **characterised in that** the returned signal (r(t)) is used as the degraded signal (s'(t)) and that an intermediate signal (Ps(f)) obtained during an intermediate stage of the second processing step of processing the reference signal is combined with a corresponding intermediate signal (Ps'(f)) obtained during a corresponding intermediate stage of the first processing step of processing the degraded signal.

8. Method according to claim 7, **characterised in that** the intermediate signal is an Fast Fourier Transform power representation (Ps(f)) of the reference speech signal (s(t)).

9. Method according to any of the claims 1,-,8, **characterised in that** the talker speech signal and the returned signal are taken off from an established telephone link.

10. Device for measuring a talking quality of a telephone link in a telecommunications network (10), the device comprising measurement means (22; 31, 36) for subjecting a degraded speech signal s'(t) with respect to a reference speech signal s(t) to an objective measurement technique for measuring a perceptual quality of speech signals, and producing a quality signal (q) which represents an estimated value concerning the talking quality, the reference speech signal being a talker speech signal (s(t)) and the degraded speech signal including a returned signal r(t), the returned signal being a signal which occurred or may occur in a return channel of the telephone link during the transmission of the talker speech signal in a forward channel of the tel-

ephone link,
characterised in that the measurement means include means (32b) for a modelling of masking effects in consequence of noise present in the returned signal.

11. Device according to claim 10, **characterised in that** the device comprises:

first processing means (39) for processing the degraded speech signal (s'(t)) and generating a first representation signal (R'(t,f)),
second processing means (38) for processing the talker speech signal (s(t)) and generating a second representation signal (R(t,f)), and
combining means (32) for combining the first and second representation signals as to produce said output signal (q), the combining means including said means (32b) for modelling the masking effects.

12. Device according to claim 11, **characterised in that**

the combining means include:

subtracting means (40) for subtracting the first representation signal from the second representation signal as to produce a difference signal (D(t,f)),
said modelling means (41, 42)) for modelling the masking effects carried out on the difference signal as to produce a modified difference signal, and
integrating means (43) for integrating the modified difference signal with respect to frequency and time as to produce the quality signal.

modelling means include:

means (41) for producing an estimated value (Ne) of the loudness of the noise present in the returned signal, and
means (42) for carrying out a noise suppression on the difference signal using said produced estimated value (Ne), and for producing the modified difference signal (D'(t,f)),

the first representation signal (R'(t,f)) being a representation signal of a signal combination of the talker speech signal and the returned signal.

12. Device according to claims 11, **characterised in that** the device includes a signal combinator (24) for combining the talker speech signal (s(t)) and the returned signal (r(t)) as to form the degraded signal (s'(t)).

13. Device according to claim 11 , **characterised in that** the device includes an intermediate signal combination means (55) for combining an intermediate signal (Ps(f)) obtained in an intermediate stage of the second processing means (38) with a corresponding intermediate signal (Ps'(f)) obtained in a corresponding intermediate stage of the first processing means (39), the degraded signal (s'(t)) being the returned signal (r(t)).

14. Device according to claim 13, **characterised in that** the intermediate signal combination means (55) is included in the first processing means (39) after means (FTT) for performing a Fast Fourier Transform.

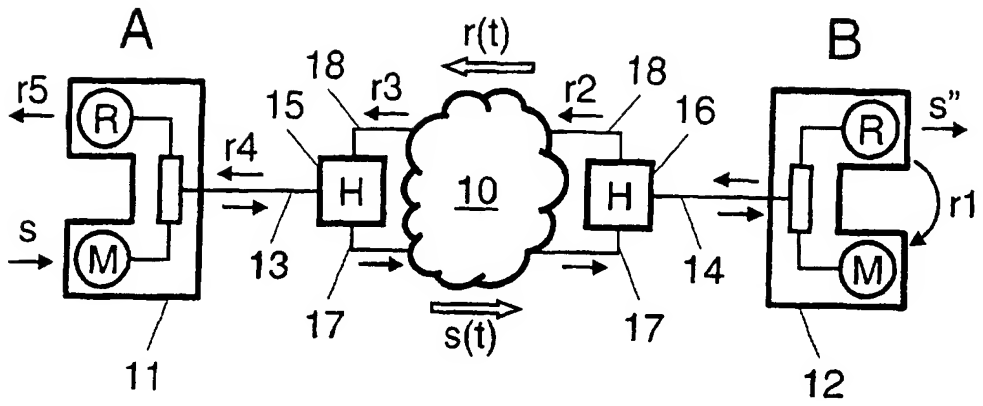
A
B
r(t)
18
r3
15
r2
18
16
r5
R
r4
H
10
H
R
s"
r1
s
M
13
14
M
17
s(t)
17
11
12

FIG. 1

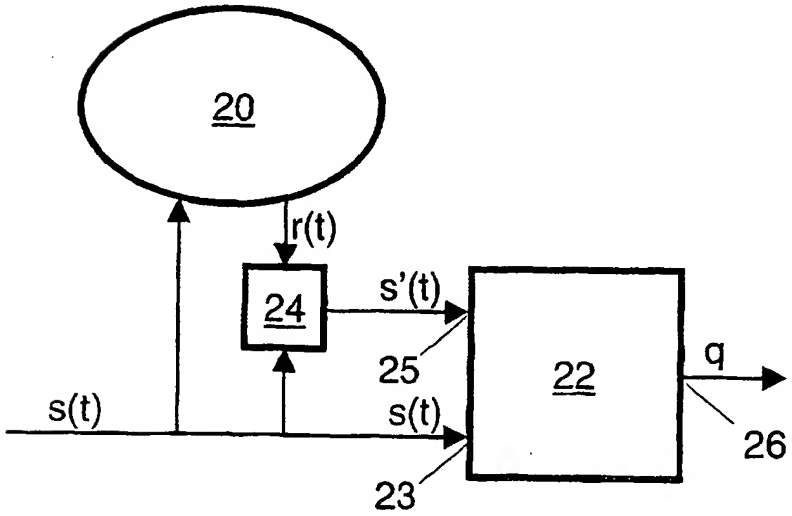
20
r(t)
24
s'(t)
25
22
q
s(t)
s(t)
26
23

FIG. 2

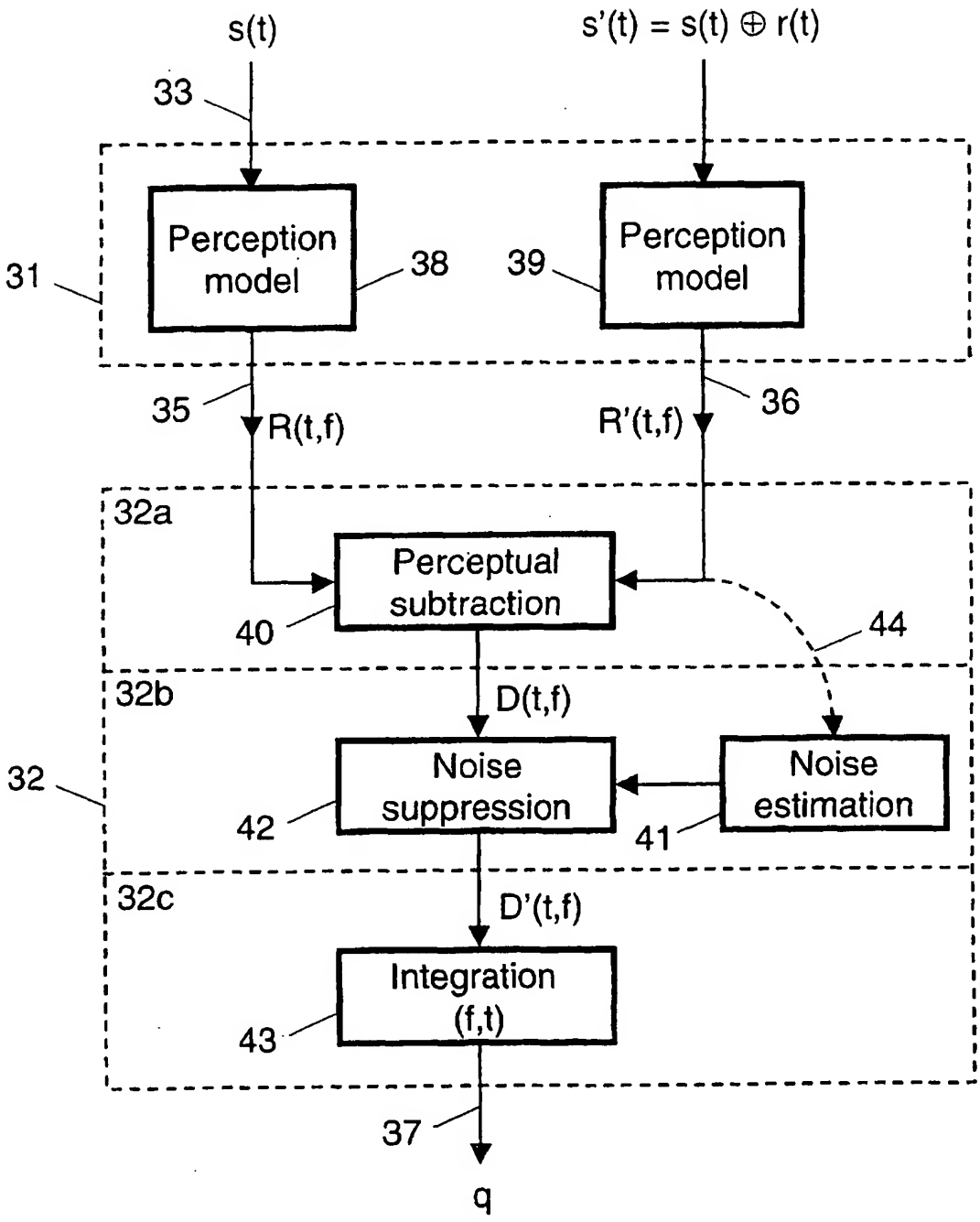
s(t)
s'(t) = s(t) ⊕ r(t)
33
Perception model
38
39
Perception model
31
35
R(t,f)
R'(t,f)
36
32a
Perceptual subtraction
40
44
32b
D(t,f)
Noise suppression
42
Noise estimation
41
32
32c
D'(t,f)
Integration (f,t)
43
37
q

FIG. 3

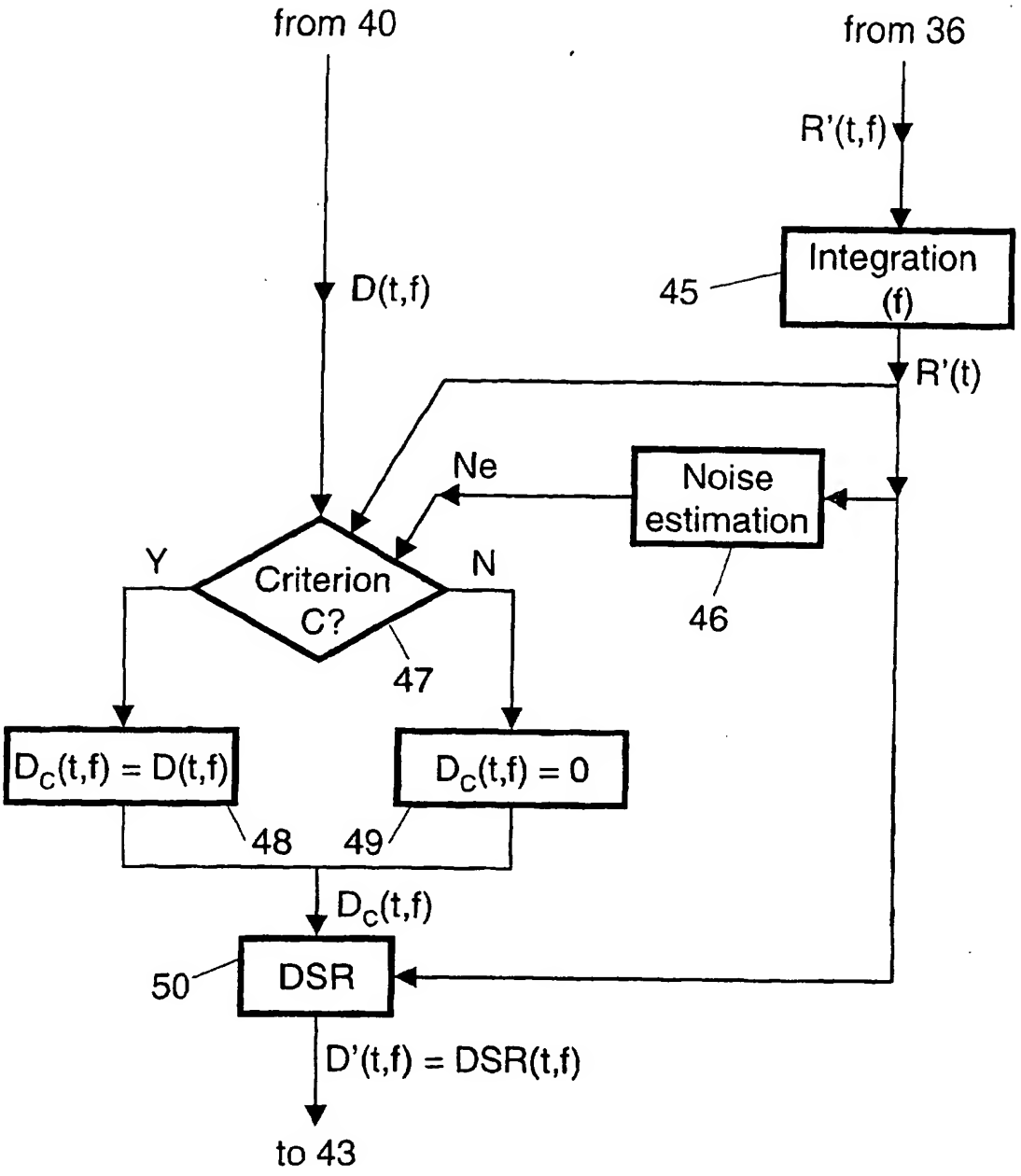
from 40
from 36
R'(t,f)
D(t,f)
45
Integration (f)
R'(t)
Ne
Noise estimation
46
Y
Criterion C?
N
47
$D_C(t,f) = D(t,f)$
$D_C(t,f) = 0$
48
49
$D_C(t,f)$
50
DSR
D'(t,f) = DSR(t,f)
to 43

FIG. 4

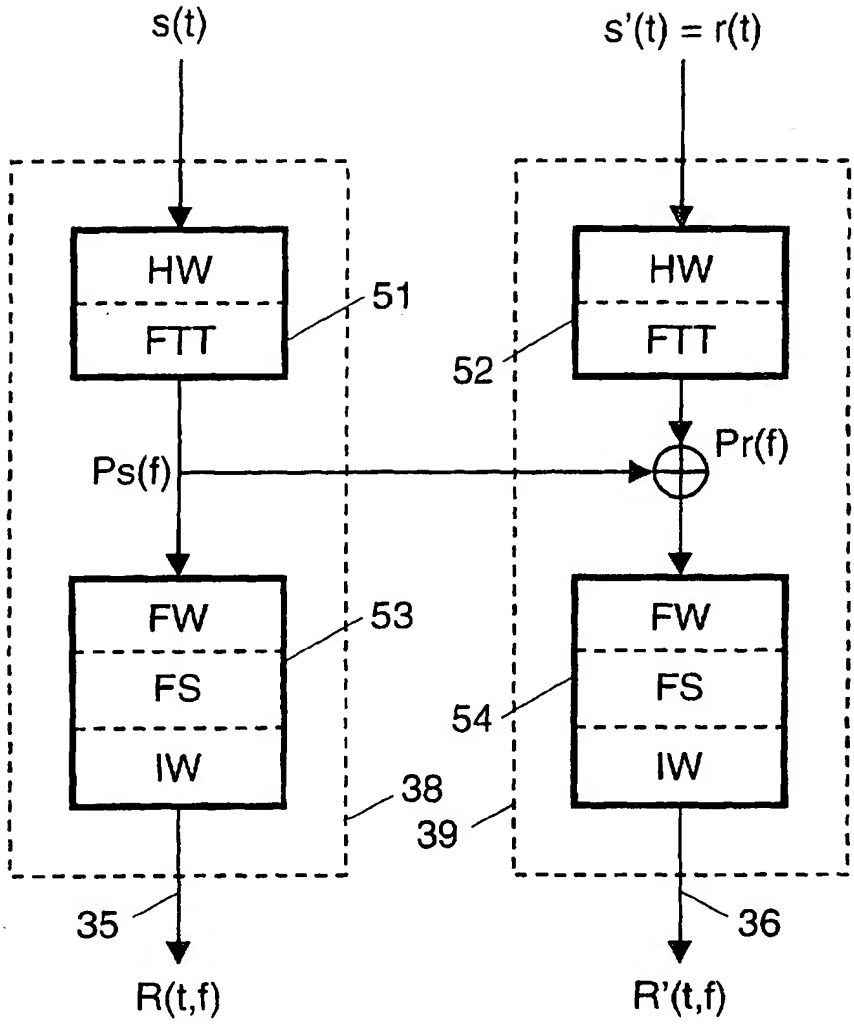
s(t)
s'(t) = r(t)
HW
FTT
51
52
HW
FTT
Ps(f)
Pr(f)
FW
FS
IW
53
54
FW
FS
IW
38
39
35
36
R(t,f)
R'(t,f)

FIG. 5

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 00 20 3936

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
Y	WO 98 59509 A (ERICSSON TELEFON AB L M) 30 December 1998 (1998-12-30) * abstract * * column 4, line 11 - column 5, line 10 * * column 6, line 2 - column 7, line 30; figure 2 * * column 8, line 20 - column 9, line 8 *	1,9,10	H04M3/32
Y	US 4 677 676 A (ERIKSSON LARRY J) 30 June 1987 (1987-06-30) * abstract *	1,9,10	
A	VASEGHI S V ET AL: "NOISE COMPENSATION METHODS FOR HIDDEN MARKOV MODEL SPEECH RECOGNITION IN ADVERSE ENVIRONMENTS" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING,US,IEEE INC. NEW YORK, vol. 5, no. 1, 1997, pages 11-21, XP000785324 ISSN: 1063-6676 the whole document	1,10	TECHNICAL FIELDS SEARCHED (Int.Cl.7) H04M

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	19 April 2001	Willems, B

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 00 20 3936

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

19-04-2001

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9859509 A	30-12-1998	US 6201960 B	13-03-2001
		AU 7950598 A	04-01-1999
		BR 9810326 A	05-09-2000
US 4677676 A	30-06-1987	AT 69660 T	15-12-1991
		AU 590384 B	02-11-1989
		AU 6860487 A	13-08-1987
		CA 1281294 A	12-03-1991
		DE 3774587 A	02-01-1992
		EP 0233717 A	26-08-1987
		ES 2028063 T	01-07-1992
		JP 2539812 B	02-10-1996
		JP 62193310 A	25-08-1987

EPO FORM P045E

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82